CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

From Genotype to Phenotype

Reconstructing the past:
The First Ancient Human Genome

Course:
IT and Health

DTU
Kasper Nielsen
Center for Biological Sequence Analysis
Technical University of Denmark

Outline

- The first ancient genome
- Functional SNP assessment
- The ancient genome database
- Genome-wide association studies
- Exercise: From genotype to phenotype

First Ancient Human Genome

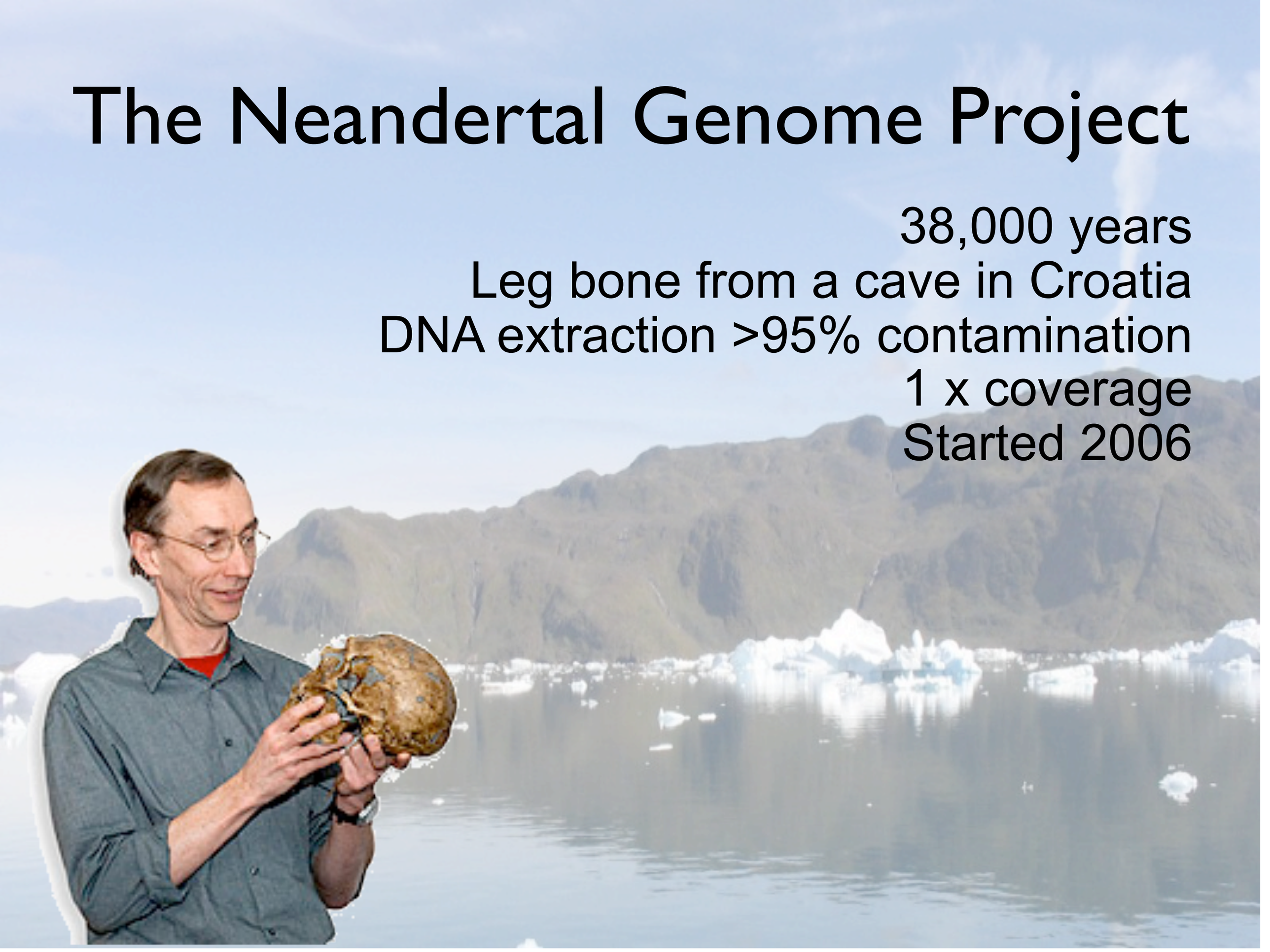
Eske Willerslev

The Saqqaq Genome Project

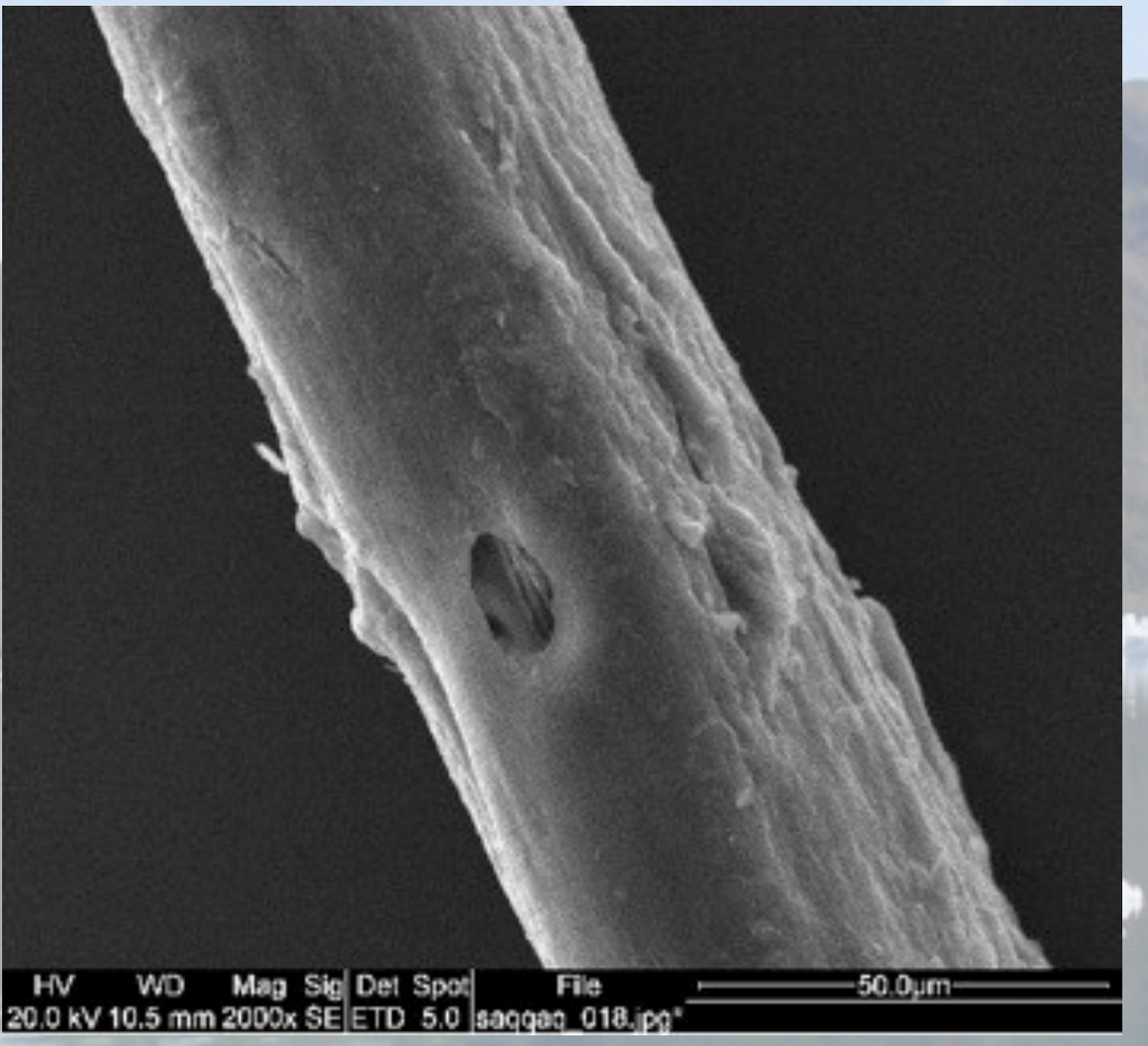
4,000 years
Hair sample from permafrost
DNA extraction <10% contamination
20 x coverage
Started 2009

The Neandertal Genome Project

38,000 years
Leg bone from a cave in Croatia
DNA extraction >95% contamination
1 x coverage
Started 2006



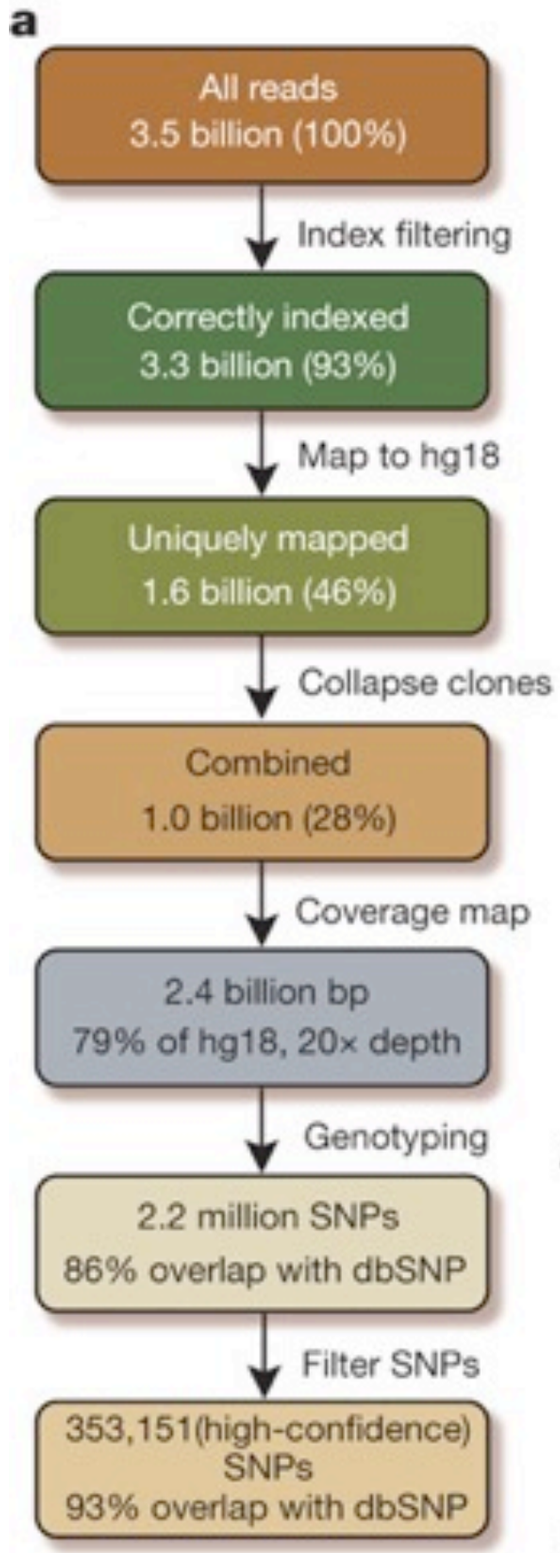
DNA from hair

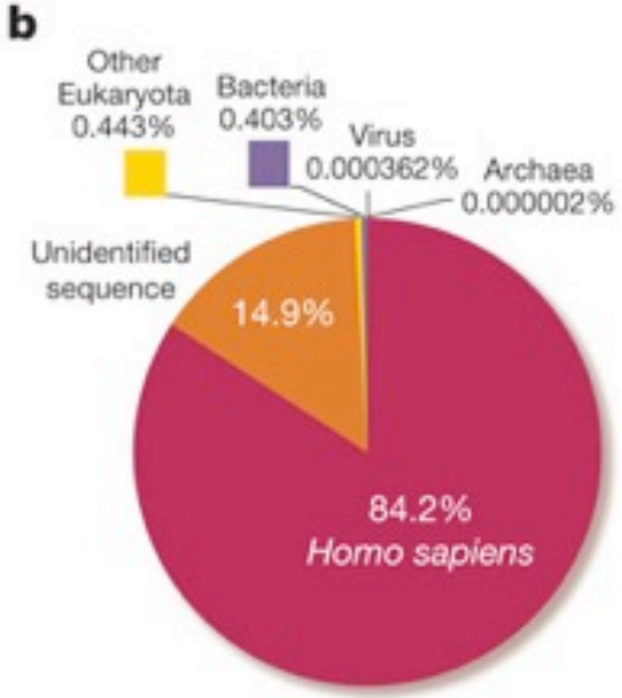


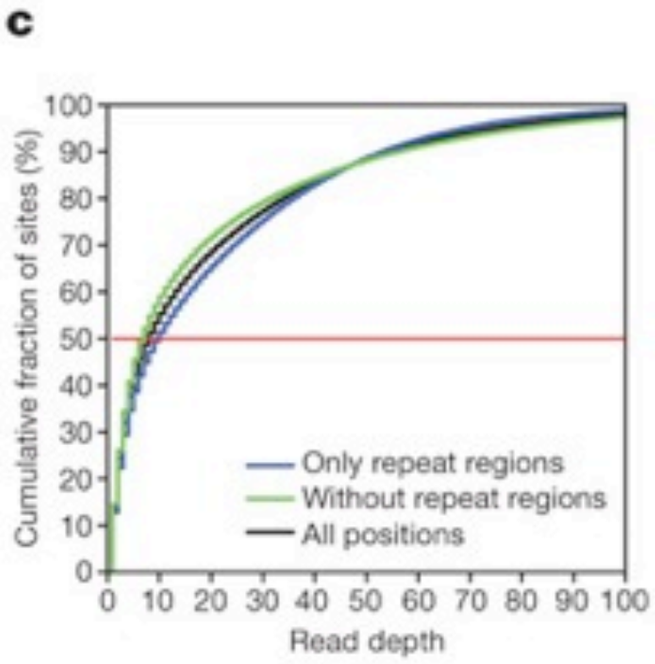
DNA from hair

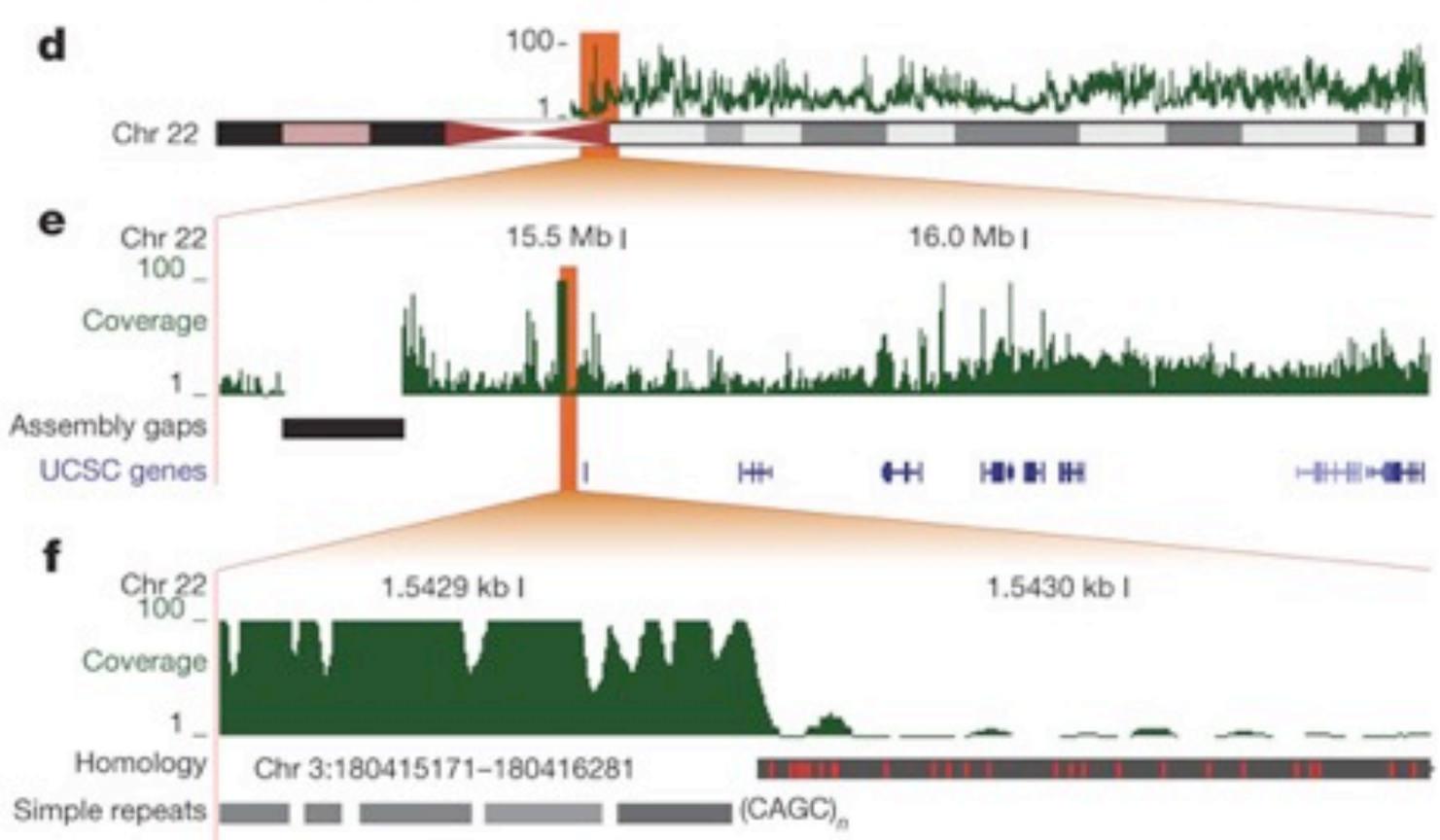
- Hardly any contamination
- Relatively high yield
- Short fragments
- <3 month sequencing
- Post-mortem DNA damage
- Polymerase Phusion

Flow chart over pipeline and data summary



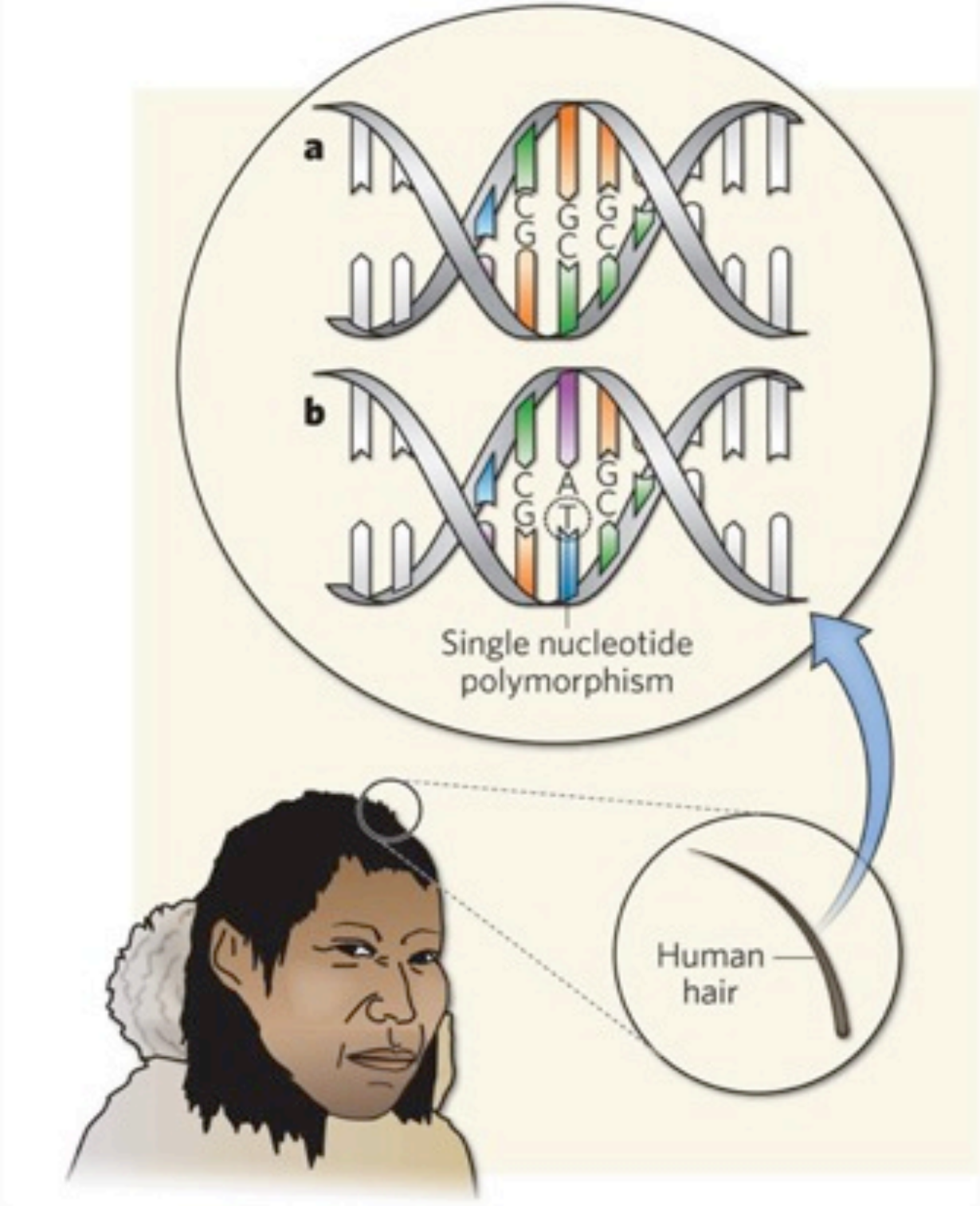






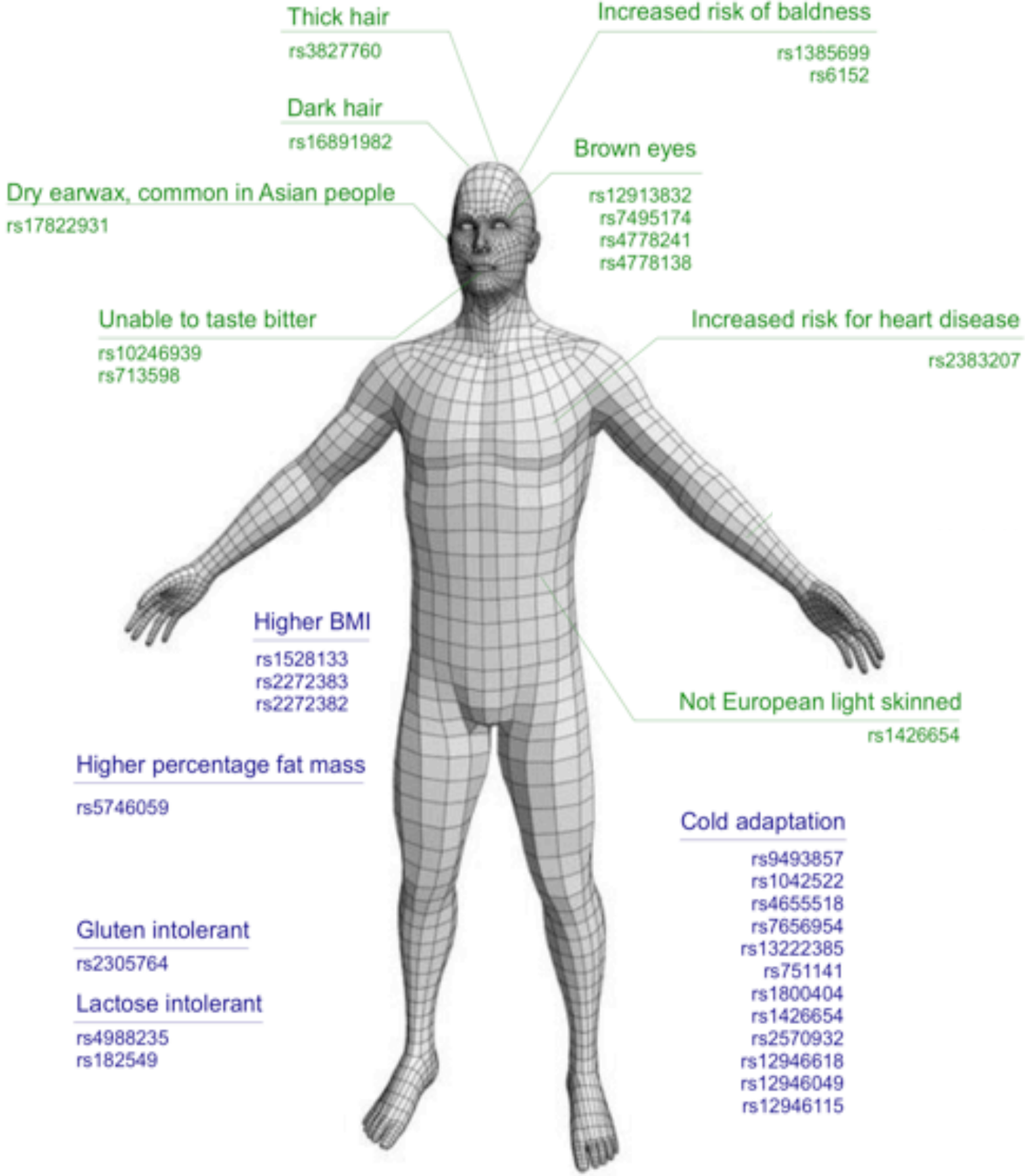
Functional SNP assessment

What can we say about his phenotype?



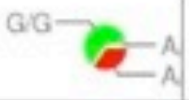
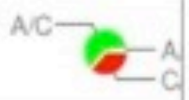
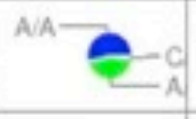
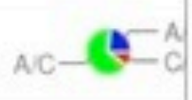
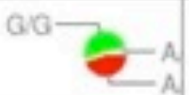
Rasmussen *et al.*[2] have sequenced the genome of a man from the Saqqaq culture, using DNA from hair preserved in permafrost in Greenland. They analysed the genome to find single nucleotide polymorphisms (SNPs) — differences in single DNA base pairs that exist between individual genomes, and that may act as markers of an individual's physical traits. **a**, Here, a short stretch of human DNA is shown that is a marker for normal earwax. **b**, In the analogous DNA from the Saqqaq individual, there is a SNP in which a C in the lower strand has been replaced by a T (C, G, T and A denote the four kinds of DNA base). This SNP shows that the Saqqaq man had dry earwax. Rasmussen and colleagues identified other SNPs indicating that the ancient human had, among other things, brown eyes, non-white skin, thick dark hair and an increased susceptibility to baldness.

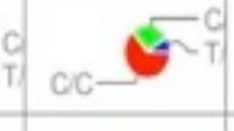
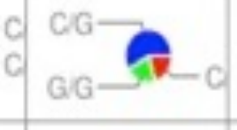
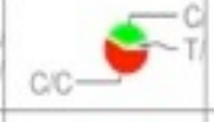
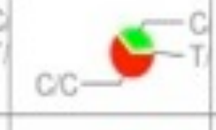
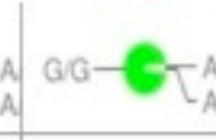
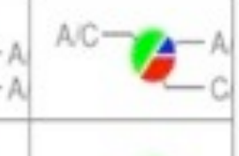
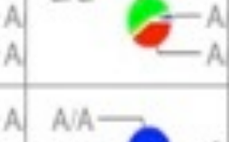
V. WARD, UNIV. AUCKLAND



From genotype to phenotype: how good are we at putting a face to an anonymous individual? - While some traits manifest themselves in a tissue specific manner (highlighted in green), others are more systemic (highlighted in blue). Going from the genetic blueprint to visual appearance, physiological behaviour and medical predispositions is still an open challenge.

Table S14. List of SNP identifiers used for functional assessment. Associations to phenotype were curated from literature, public databases and HGMD Professional. Frequencies from Hapmap Phase-II are represented in pie charts, with the colour of the Saqqaq Allele matching the corresponding genotype in the pie charts. In most cases, the Saqqaq allele is closest to the Asian (Han Chinese and Japanese) populations.

SNP	Reference	1-pp	Saqqaq Allele	HapMap frequencies			
				CEU	*HCB*	*JPT*	*YRI*
rs8176719[1]	66	NA	-/-	NA	NA	NA	NA
rs8176750[2]	67	8.84E-34	G/G	NA	NA	NA	NA
rs12913832[3]	68	5.13E-10	A/A				
rs7495174[3]	68	1.14E-25	A/G				
rs4778241[3]	68	8.24E-06	A/A				
rs1129038[3]	68	8.69E-40	C/C	NA	NA	NA	NA
rs1426654[4]	69	8.78E-12	G/G				
rs1385699[5]	70	8.57E-05	T/T				
rs6152[5]	71	4.12E-21	G/G	NA	NA	NA	NA
rs1528133[6]	63	1.50E-08	A/C				
rs2272383[6]	63	1.42E-27	A/G				
rs2272382[6]	63	8.40E-60	A/G				
rs5746059[7]	62	1.64E-05	A/A				

rs17822931[8]	72	8.36E-09	T/T				
rs3827760[9]	73	5.61E-12	C/C				
rs16891982[10]	74	3.87E-13	C/C				
rs1042522[11]	75	2.75E-11	G/G				
rs13222385[12]	59	3.21E-05	A/G				
rs751141[12]	59	2.13E-07	T/T				
rs1800404[12]	59	2.51E-14	A/A				
rs1426654[12]	59	8.78E-12	G/G				
rs2570932[12]	59	0.0015097	C/C				
rs12946618[12]	59	0.0013802	A/A				
rs12946115[12]	59	2.44E-06	C/C				

[1]Blood Group: not type O. [2]Blood Group: A1 subtype. [3]Brown eyes. [4]Not European light skinned. [5]Increased risk of baldness. [6]Higher body mass index. [7]Higher percentage fat mass in Caucasian and Chinese samples. [8]Dry earwax, common in Asian people. [9]Thick hair and Shovel shaped upper front teeth. [10]More likely to have black hair (in European cohort study). [11] Cold adaptation: non-synonymous change in TP53. [12] Cold adaptation: Metabolic genes.

The ancient genome database

www.ancientgenome.dk

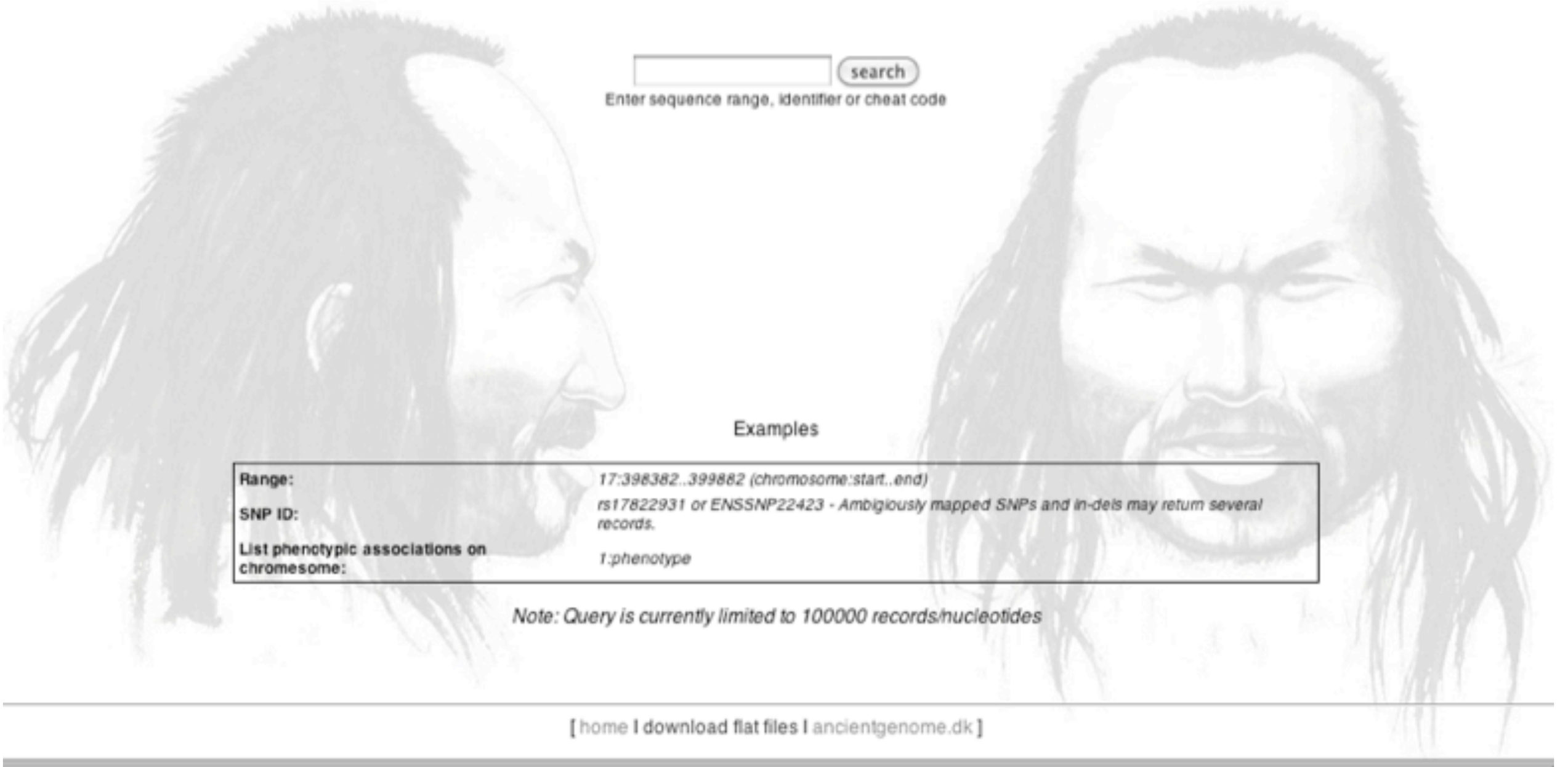
- Get sequence
- Compare to reference genome (NCBI 36)
- SNP accessions
- Phenotypic associations

The Saqqaq Genome Database (beta)

http://www.cbs.dtu.dk/cgi-bin/saqqaq.cgi

Google

The Saqqaq Genome Database (beta)



search

Enter sequence range, identifier or cheat code

Examples

Range:	*17:398382..399882 (chromosome:start..end)*
SNP ID:	*rs17822931 or ENSSNP22423 - Ambigiously mapped SNPs and in-dels may return several records.*
List phenotypic associations on chromesome:	*1:phenotype*

Note: Query is currently limited to 100000 records/nucleotides

The Saqqaq Genome Database (beta)

Choose alternative output format: Table with additional information Submit

Result

chr	pos	ref	is_ref	genotype	pp	depth	repeat	rs	type	strand	snp_alleles
16	46815689	C	y	CC	1.6396e-13	33					
16	46815690	T	y	TT	9.9824e-14	33					
16	46815691	T	y	TT	9.995e-14	33					
16	46815692	A	y	AA	1.01e-13	33					
16	46815693	C	y	CC	3.2608e-13	32					
16	46815694	T	y	TT	1.9992e-13	32					
16	46815695	G	y	GG	6.565e-13	31					
16	46815696	G	y	GG	6.9136e-13	31					
16	46815697	C	y	CC	1.3222e-12	30					
16	46815698	C	y	CC	2.6699e-12	29					
16	46815699	C	n	TT	8.3635e-09	28		rs17822931	single	+	CT
16	46815700	G	y	GG	1.3271e-12	30					
16	46815701	A	y	AA	2.5557e-13	31					
16	46815702	G	y	GG	2.0907e-13	32					
16	46815703	T	y	TT	6.4711e-14	33					
16	46815704	A	y	AA	1.3049e-13	32					
16	46815705	C	y	CC	4.214e-13	31					
16	46815706	A	y	AA	2.5604e-13	31					
16	46815707	C	y	CC	4.4147e-13	31					
16	46815708	T	y	TT	2.5521e-13	31					
16	46815709	G	y	GG	2.1138e-13	32					

Columns explained

chr	The chromosome
pos	Position on chromosome
ref	The reference nucleotide on leading strand in hg18
is_ref	Indicates whether the genotype is the same as the reference nucleotide (y) or not (n)
genotype	The genotype called (on leading strand)
pp	For numerical reasons, we report (1-PP), where PP is the posterior probability of the genotype
depth	The number of reads covering the position
repeat	If the position lies in an annotated repeat, the ID is given here
rs	dbSNP rs-number
type	Type of dbSNP entry ("single", "indel" etc. - see the UCSC genome browser for details).
strand	Strand for dbSNP entry + (or 1) or - (or -1)
snp_alleles	Known SNP alleles, e.g. "AC" (or "A/G") for a SNP of type "single"

Genome-wide associations studies

Cases vs controls

- Obtain DNA from a disease group (e.g. asthma) and a control group
- For each individual: Run DNA on a SNP array measuring the genotypes at almost 1M loci
- If the group consists of 1,000 individuals, one obtains 1,000 x 1,000,000 genotypes
- Quality Control
- Loop over all SNPs, and identify those that are significantly more common among cases than controls
- Those SNPs are associated with the disease
- Not necessarily causal

Example of GWAS results (asthma)

Manhattan plot displays all SNPs on x-axis (order by genomic location), and -$\log_{10}$ of p-values on y-axis.

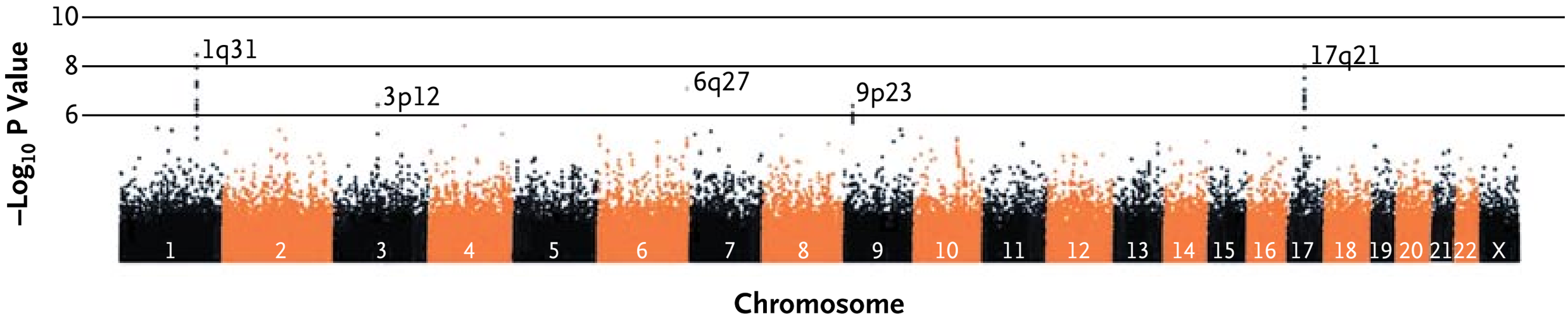


Figure 2. Manhattan Plot of the Results from the Combined Subjects of European Ancestry Who Had Asthma.

The –$\log_{10}$ P values are plotted against the physical distance. Only the two loci at chromosome 1q31 and 17q21 were significantly associated with asthma after Bonferroni correction. Individual chromosome labels are indicated in white within the Manhattan plot.

Variants of *DENND1B* Associated with Asthma in Children

Patrick M.A. Sleiman, Ph.D., James Flory, Ph.D., Marcin Imielinski, M.D., Ph.D., Jonathan P. Bradfield, B.S., Kiran Annaiah, M.Sc., Saffron A.G. Willis-Owen, Ph.D., Kai Wang, Ph.D., Nicholas M. Rafaels, M.S., Sven Michel, Ph.D., Klaus Bonnelykke, M.D., Ph.D., Haitao Zhang, Ph.D., Cecilia E. Kim, B.A., Edward C. Frackelton, B.A., Joseph T. Glessner, M.Sc., Cuiping Hou, M.Sc., F. George Otieno, M.Sc., Erin Santa, B.A., Kelly Thomas, B.A., Ryan M. Smith, B.A., Wendy R. Glaberson, B.A., Maria Garris, B.A., Rosetta M. Chiavacci, B.S.N., Terri H. Beaty, Ph.D., Ingo Ruczinski, Ph.D., Jordan M. Orange, M.D., Ph.D., Julian Allen, M.D., Jonathan M. Spergel, M.D., Ph.D., Robert Grundmeier, M.D., Ph.D., Rasika A. Mathias, Sc.D., Jason D. Christie, M.D., Erika von Mutius, M.D., William O.C. Cookson, M.D., Michael Kabesch, M.D., Miriam F. Moffatt, Ph.D., Michael M. Grunstein, M.D., Ph.D., Kathleen C. Barnes, Ph.D., Marcella Devoto, Ph.D., Mark Magnusson, M.D., Hongzhe Li, Ph.D., Struan F.A. Grant, Ph.D., Hans Bisgaard, M.D., and Hakon Hakonarson, M.D., Ph.D.

N Engl J Med 2010; 362:36-44 | January 7, 2010

Online phenotype association resources

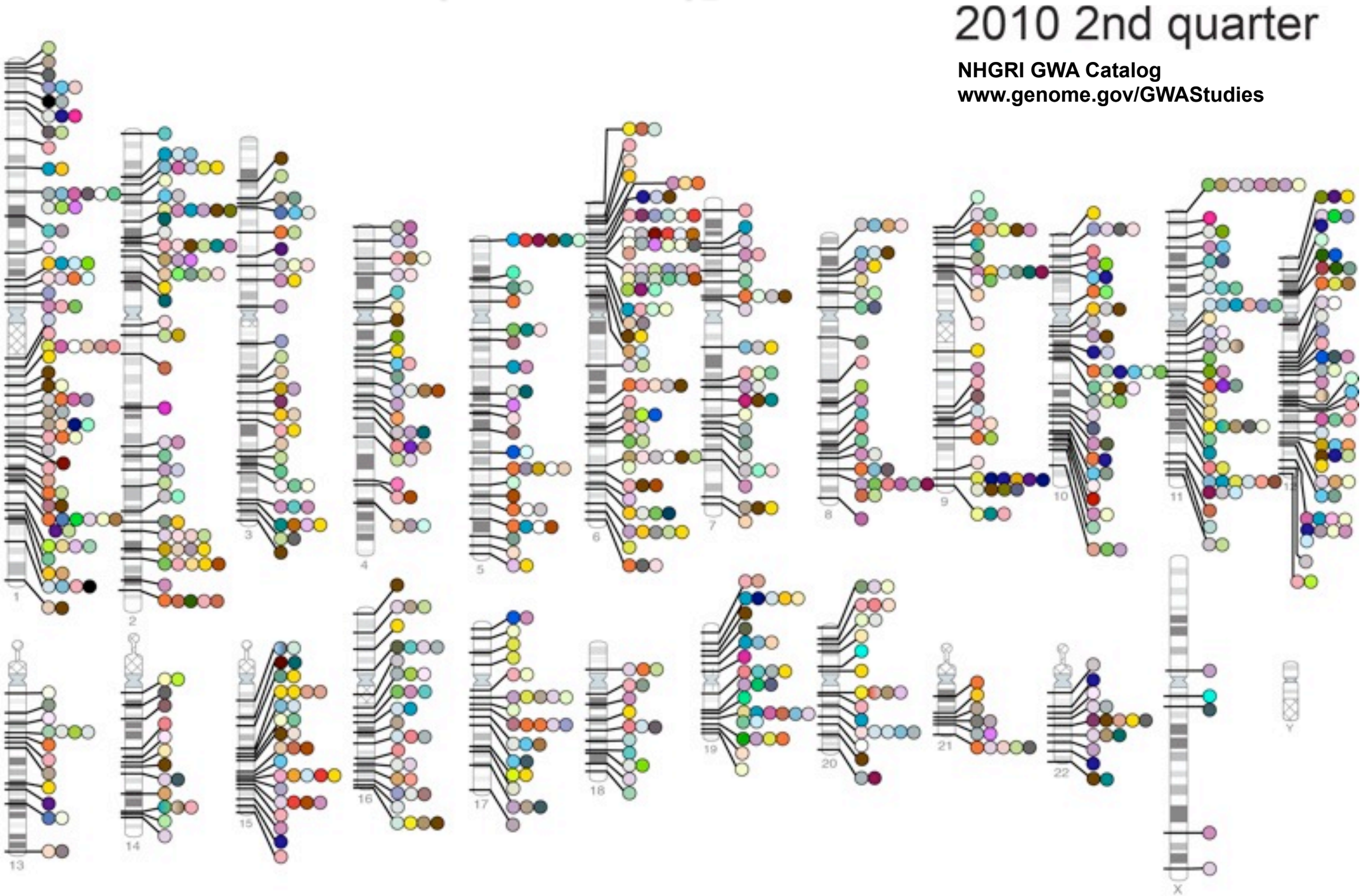
- NHGRI's Catalog of Published GWAS:

  http://www.genome.gov/26525384

- SNPedia:

  http://www.snpedia.com

Published Genome-Wide Associations through 6/2010, 904 published GWA at $p \leq 5 \times 10^{-8}$ for 165 traits



- Acute lymphoblastic leukemia
- Adhesion molecules
- Adiponectin levels
- Age-related macular degeneration
- AIDS progression
- Alcohol dependence
- Alzheimer disease
- Amyotrophic lateral sclerosis
- Angiotensin-converting enzyme activity
- Ankylosing spondylitis
- Arterial stiffness
- Asthma
- Atherosclerosis in HIV
- Atrial fibrillation
- Attention deficit hyperactivity disorder
- Autism
- Basal cell cancer
- Bipolar disorder
- Biliary atresia
- Bilirubin
- Birth weight
- Bladder cancer
- Blond or brown hair
- Blood pressure
- Blue or green eyes
- BMI, waist circumference
- Bone density
- Breast cancer
- C-reactive protein
- Cardiac structure/function
- Carnitine levels
- Carotenoid/tocopherol levels
- Celiac disease
- Chronic lymphocytic leukemia
- Cleft lip/palate
- Cognitive function
- Conduct disorder
- Colorectal cancer
- Corneal thickness
- Coronary disease
- Creutzfeldt-Jakob disease
- Crohn's disease
- Cutaneous nevi
- Dermatitis
- Drug-induced liver injury
- Eosinophil count
- Eosinophilic esophagitis
- Erythrocyte parameters
- Esophageal cancer
- Essential tremor
- Exfoliation glaucoma
- Eye color traits
- F cell distribution
- Fibrinogen levels
- Folate pathway vitamins
- Freckles and burning
- Gallstones
- Glioma
- Glycemic traits
- Hair color
- Hair morphology
- HDL cholesterol
- Heart failure
- Heart rate
- Height
- Hemostasis parameters
- Hepatitis
- Hirschsprung's disease
- HIV-1 control
- Homocysteine levels
- Idiopathic pulmonary fibrosis
- IgE levels
- Inflammatory bowel disease
- Intracranial aneurysm
- Iris color
- Iron status markers
- Ischemic stroke
- Juvenile idiopathic arthritis
- Kidney stones
- LDL cholesterol
- Leprosy
- Leptin receptor levels
- Liver enzymes
- LP (a) levels
- LpPLA(2) activity and mass
- Lung cancer
- Major mood disorders
- Malaria
- Male pattern baldness
- Matrix metalloproteinase levels
- MCP-1
- Melanoma
- Menarche & menopause
- Multiple sclerosis
- Myeloproliferative neoplasms
- Narcolepsy
- Nasopharyngeal cancer
- Neuroblastoma
- Nicotine dependence
- Obesity
- Open angle glaucoma
- Open personality
- Optic disc parameters
- Osteoarthritis
- Osteoporosis
- Otosclerosis
- Other metabolic traits
- Ovarian cancer
- Pancreatic cancer
- Pain
- Paget's disease
- Panic disorder
- Parkinson's disease
- Periodontitis
- Peripheral arterial disease
- Phosphatidylcholine levels
- Phytosterol levels
- Platelet count
- Primary biliary cirrhosis
- PR interval
- Prostate cancer
- Protein levels
- Psoriasis
- Pulmonary funct. COPD
- QRS interval
- QT interval
- Quantitative traits
- Recombination rate
- Red vs.non-red hair
- Renal function
- Response to antidepressants
- Response to antipsychotic therapy
- Response to hepatitis C treat
- Response to statin therapy
- Restless legs syndrome
- Rheumatoid arthritis
- Schizophrenia
- Serum metabolites
- Skin pigmentation
- Smoking behavior
- Speech perception
- Sphingolipid levels
- Statin-induced myopathy
- Stroke
- Systemic lupus erythematosus
- Systemic sclerosis
- Telomere length
- Testicular germ cell tumor
- Thyroid cancer
- Tooth development
- Total cholesterol
- Triglycerides
- Type 1 diabetes
- Type 2 diabetes
- Ulcerative colitis
- Urate
- Venous thromboembolism
- Vertical cup-disc ratio
- Vitamin B12 levels
- Vitamin D insuffiency
- Vitiligo
- Warfarin dose
- Weight
- White cell count
- YKL-40 levels

SNPedia

Earwax

rs17822931 determines wet vs dry earwax [PMID 16444273]

This can also be used to distinguish asian ancestry.

NCBI coffeebreak introduction

Category: Is a medical condition

Navigation

SNPedia
FAQ
Blog
Recent changes
Random page

Toolbox

What links here
Related changes
Special pages
Printable version
Permanent link
Browse properties

From genotype to phenotype

- 99.9% of human DNA is identical to another person
- 0.1% → over 80% are single nucleotide polymorphisms (SNPs)
- The genotype of an organism is the inherited instructions within the DNA
- A phenotype is any observable characteristic or trait of an organism

Exercise: Genotype to Phenotype

http://wiki.cbs.dtu.dk/teachingmaterials/index.php/ExGenotype2PhenotypeLite